Language is funny

"Red tape holds up new bridges"

"Hospitals are sued by 7 foot doctors"

"Local high school dropouts cut in half"

"Tesla crashed today"

"Obama announced that he will run again"

"Kipchoge announced that he will run again"

"She made him duck"

"Will you visit the bank across from the river bank? You can bank on it"

"Yes" vs "Yes." vs "YES" vs "YES!" vs "YAS" vs "Yea"

Multiple levels* to a single word

Discourse

Pragmatics

Semantics

Syntax

Lexemes

Morphology

phonology

phonetics

speech

orthography

text

- Inputs (words) are noisy
- Capture theoretical concepts that are ~latent variables
- Ambiguity abound. Many interpretations at each level

*

Slide adapted from or inspired by Chris Manning and Richard Socher

Common NLP Tasks (aka problems)

Syntax

Morphology

Word Segmentation

Part-of-Speech Tagging

Parsing

- Constituency
- Dependency

Discourse

Summarization

Coreference Resolution

Semantics

Sentiment Analysis

Topic Modelling

Named Entity Recognition (NER)

Relation Extraction

Word Sense Disambiguation

Natural Language Understanding (NLU)

Natural Language Generation (NLG)

Machine Translation

Entailment

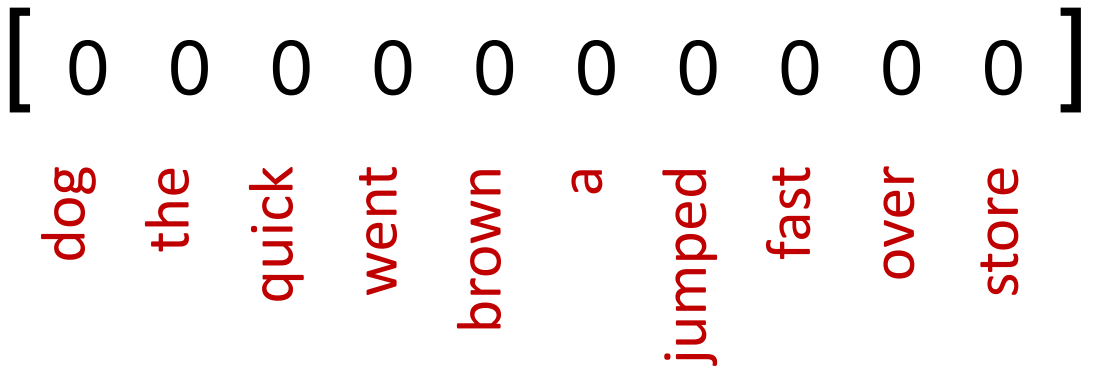
Question Answering

Language Modelling

Feature Vector

Let's say our dataset's entire *vocabulary* is just 10 words.

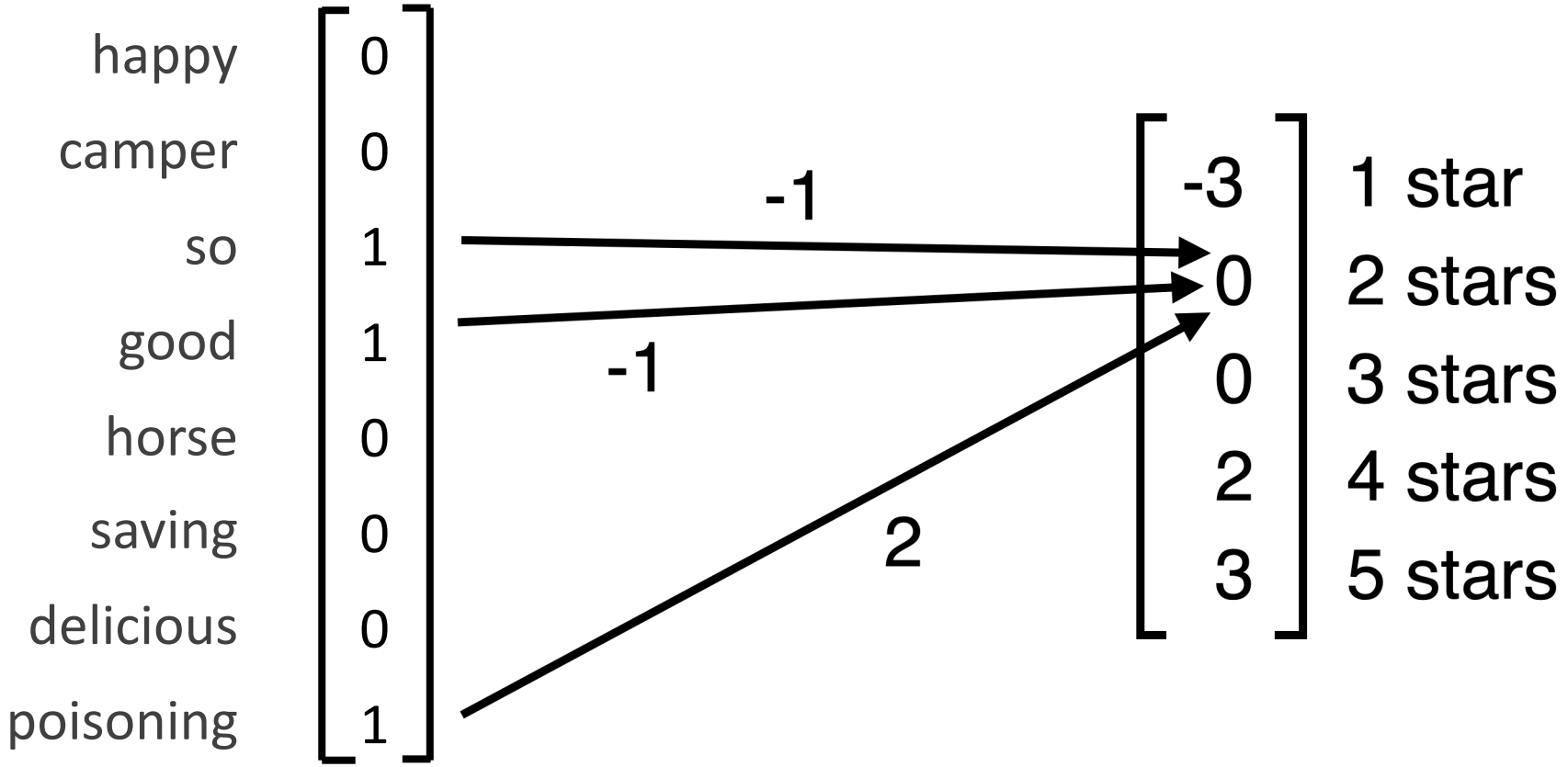
Each unique word can have its own dimension (feature index).



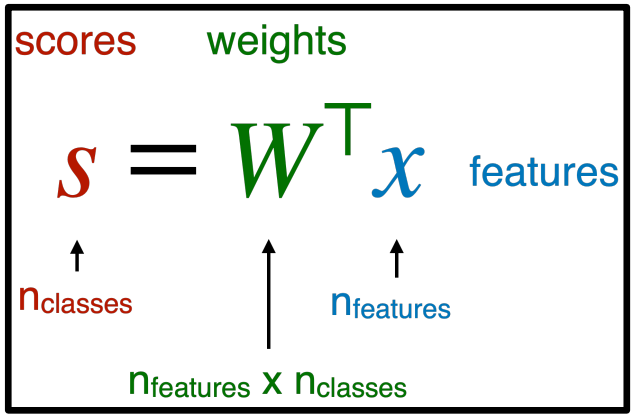
NOTE: This is the Boolean version of bag-of-words (BoW), which isn't the most popular BoW representation

Linear Models

"So good, but gave me food poisoning"



Linear Models



Loss: Negative Log Likelihood

$$L(s, y) = -\log p(y \mid x)$$

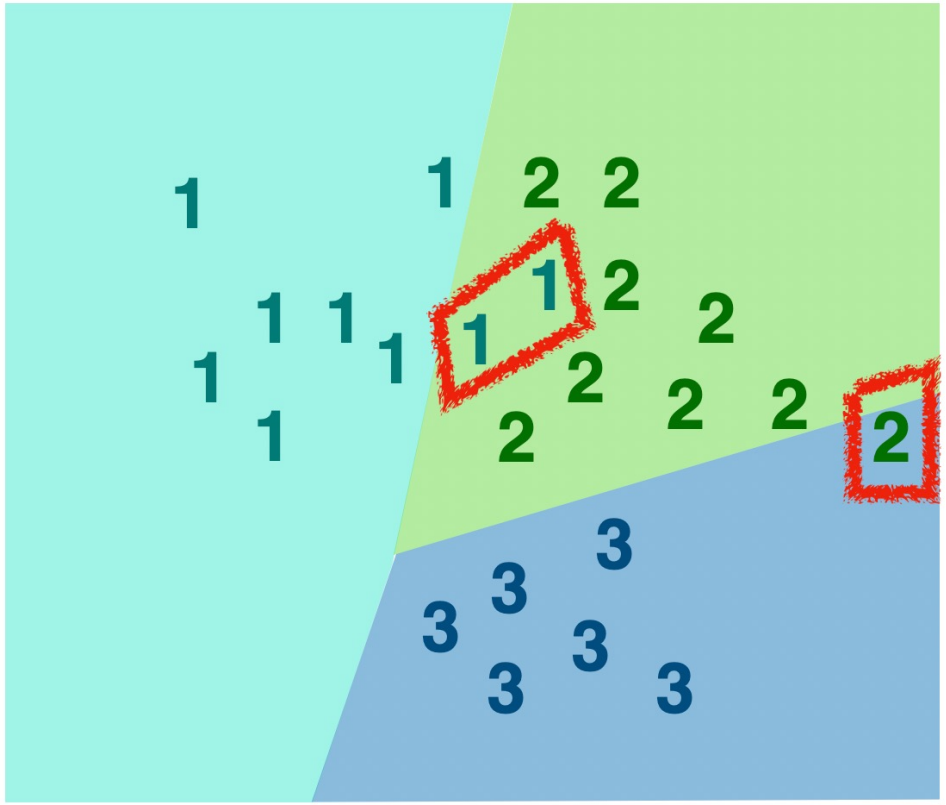
$$= -\log \frac{\exp(s_y)}{\sum_i \exp(s_i)}$$

$$= -s_y + \log \sum_i \exp(s_i) := -\log \text{softmax}(s)_y$$

Idea: treat s as a vector of (unnormalized) log-probs, and maximize p(y | x; W).

Non-Linear Models

Our old friend, k-NN



Non-Linear Models

Deep Networks

Linear model: $s = W^\top x$

Deep linear model: $s = W_2^\top W_1^\top$

(same expressive power!)

Neural network model: $s = W_2^\top f(W_1^\top x)$